Adaptive Information Selection in Images: Efficient Naive Bayes Nearest Neighbor Classification

Thomas Reineking(✉), Tobias Kluth, and David Nakath

Cognitive Neuroinformatics, University of Bremen,
Enrique-Schmidt-Str. 5, 28359 Bremen, Germany
{trking,tkluth,dnakath}@cs.uni-bremen.de

Abstract. We propose different methods for adaptively selecting information in images during object recognition. In contrast to standard feature selection, we consider this problem in a Bayesian framework where features are sequentially selected based on the current belief distribution over object classes. We define three different selection criteria and provide efficient Monte Carlo algorithms for the selection. In particular, we extend the successful Naive Bayes Nearest Neighbor (NBNN) classification approach, which is very costly to compute in its original form. We show that the proposed information selection methods result in a significant speed-up because only a small number of features needs to be extracted for accurate classification. In addition to adaptive methods based on the current belief distribution, we also consider image-based selection methods and we evaluate the performance of the different methods on a standard object recognition data set.

Keywords: Object recognition · Classification · Information selection · Bayesian inference · Information gain

1 Introduction

Selecting relevant information from a high-dimensional input is a fundamental problem pertaining many different areas ranging from computer vision to robotics. An effective selection strategy uses only a small subset of the available information without negatively impacting the task performance. An example of a successful selection strategy is the processing of visual information in humans where eye movements are performed in order to extract the relevant information from a scene in a very efficient manner [11]. A key feature of this selection is its adaptivity because the selection is strongly influenced by the current belief about the scene [17].

In this paper, we follow the idea of an adaptive belief-based information selection and we investigate it in the context of object recognition. While object recognition is usually viewed as a static pattern recognition problem, we model the recognition as an information gathering process unfolding in time, which is

G. Azzopardi and N. Petkov (Eds.): CAIP 2015, Part I, LNCS 9256, pp. 350–361, 2015.
DOI: 10.1007/978-3-319-23192-1_29

more akin to visual processing in humans. In this case, recognition becomes a problem of Bayesian information fusion where the selection of relevant information is done adaptively with regard to the current belief distribution (in contrast to classical feature selection methods, e.g. [4,7]). We propose different criteria for optimal information selection and provide efficient algorithms for their application. In addition to belief-based selection methods, we also consider an image-based method that uses a saliency operator to identify relevant locations in an image.

We combine the information selection methods with the successful NBNN object recognition approach presented in [1]. We use NBNN because it is a probabilistic approach where local image features are sequentially processed in order to update a belief distribution over possible object classes.[1] For each extracted feature, multiple expensive nearest neighbor searches have to be performed, which is why selecting a small subset of relevant features greatly reduces the computational costs of NBNN classification (for making the nearest neighbor search itself more efficient, see [9]). Note that while we focus on object recognition in this paper, the proposed belief-based information selection methods are very versatile and could therefore also be applied in other contexts.

The paper is structured as follows. In the next section, the basics of the NBNN approach are introduced. In Sect. 3, the information selection methods are described in detail. In Sect. 4, the different selection methods are combined with the NBNN approach and compared empirically on a standard object recognition data set. The paper concludes with a short discussion of the proposed methods and possible extensions.

2 Naive Bayes Nearest Neighbor

For NBNN, a set of local image descriptors is extracted from the query image (e.g. SIFT descriptors [8]) which is then used to compute the posterior probability distribution over object classes. Let $\mathcal{C}$ denote the set of object classes, and let $d_{1:N}$ denote all descriptors extracted from the query image[2] where N is the total number of descriptors found in the image. By applying Bayes' rule and by making a naive Bayes assumption regarding the conditional independence of descriptors, the posterior is given by

$$P(c|d_{1:N}) \propto P(c) \prod_{i=1}^{N} p(d_i|c) \text{ with } c \in \mathcal{C}. \tag{1}$$

The likelihood $p(d_i|c)$ for the i-th descriptor is approximated using kernel density estimation (KDE). This avoids the severe errors caused by quantizing descriptors like in bag-of-words models [2]. To reduce computational complexity and in contrast to typical KDE, only the nearest neighbor (NN) of d_i in the

[1] Other state-of-the-art classification approaches like deep networks [6] are not suited here because they do not allow for an incremental processing of features.

[2] We use the shorthand notation $d_{1:N} = d_1, \ldots, d_N$.

training set is considered because the density contributions of descriptors that are farther away tend to be negligible. Using a Gaussian kernel, the likelihood is approximated by

$$p(d_i|c) = \frac{1}{|\mathcal{D}_c|} \sum_{d^{(j)} \in \mathcal{D}_c} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{||d_i - d^{(j)}||^2}{2\sigma^2}) \tag{2}$$

$$\approx \frac{1}{\sqrt{2\pi}\sigma|\mathcal{D}_c|} \exp(-\frac{||d_i - NN_c(d_i)||^2}{2\sigma^2}) \tag{3}$$

with

$$NN_c(d_i) = \arg\min_{d^{(j)} \in \mathcal{D}_c} ||d_i - d^{(j)}|| \tag{4}$$

where σ denotes the (class-independent) KDE bandwidth, $\mathcal{D}_c$ denotes the set of descriptors in the training set belonging to class c, and $NN_c(d_i)$ denotes the NN of d_i in $\mathcal{D}_c$. The posterior is thus given by

$$P(c|d_{1:N}) \propto P(c) \prod_{i=1}^{N} p(d_i|c) \propto P(c) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N} ||d_i - NN_c(d_i)||^2\right). \tag{5}$$

Note that we ignore the descriptor count $|\mathcal{D}_c|$ for the posterior because its influence is very limited and it simplifies the derivations below. Assuming a uniform class prior, the most probable class c^* can be found using the simple decision rule

$$c^* = \arg\max_{c \in \mathcal{C}} \log P(c|d_{1:N}) = \arg\min_{c \in \mathcal{C}} \sum_{i=1}^{N} ||d_i - NN_c(d_i)||^2. \tag{6}$$

Though the decision rule in Eq. (6) is independent of σ (it is therefore ignored in the original NBNN approach), the bandwidth turns out to be relevant for the selection of optimal descriptors in the next section. We determine the optimal bandwidth σ^* by maximizing the log-likelihood of all training set descriptors $\mathcal{D} = \cup_{c \in \mathcal{C}} \mathcal{D}_c$ according to

$$\sigma^* = \arg\max_{\sigma} \log p(\mathcal{D}|\sigma) = \sqrt{\frac{\sum_{c \in \mathcal{C}} \sum_{d^{(i)} \in \mathcal{D}_c} ||d^{(i)} - NN_c(d^{(i)})||^2}{|\mathcal{D}|}}. \tag{7}$$

3 Information Selection

For selecting the most relevant descriptors, we distinguish between belief-based selection methods and image-based ones. For belief-based selection, the probabilistic model introduced in the previous section is used to predict the effect of extracting a descriptor at a particular location in the image on the current belief distribution. In contrast, for image-based selection, the image information itself is used to determine which regions in the image are most relevant without considering the training data.

We model the information selection problem as one of finding the most promising *absolute* location in an image where the object is assumed to be depicted at the center of the image. This simplification allows us to ignore the problem of object detection, which would be necessary in case of more complex scenes with variable object locations. Let l_t denote the location of a descriptor d_{l_t} in an image at the t-th extraction step after already having extracted the first $t-1$ descriptors $d_{l_1:l_{t-1}}$. To select the next optimal location, we compute a score $S(l_t)$ for each location and choose the maximum

$$l_t^* = \arg\max_{l_t \in \mathcal{L}_t} S(l_t). \tag{8}$$

To limit the number of locations, we put a grid over each image where a location represents a grid cell. Because of the naive Bayes assumption, the likelihoods of the descriptors within a cell can simply be combined by multiplying them, i.e., each likelihood $p(d_{l_t}|c)$ represents a product of the likelihoods of individual descriptors located within the same grid cell.

In the remainder of this section, we first present two belief-based information selection methods and then an image-based one.

3.1 Maximum Expected Probability

For classification it is useful to select the descriptor that maximizes the expected posterior probability (MEP) of the true class. Because the value of the next descriptor is unknown prior to extracting it, it has to be modeled as a random variable D_{l_t}. The same applies to the value of the true object class of the query image, which is modeled as a random variable $C_{\text{true}} \in \mathcal{C}$. The score S_{MEP} is the conditional expectation of the true class posterior probability

$$\begin{aligned} S_{\text{MEP}}(l_t) &= E[P(C_{\text{true}}|d_{l_1:l_{t-1}}, D_{l_t})|d_{l_1:l_{t-1}}] && (9)\\ &= \int \sum_{c_{\text{true}} \in \mathcal{C}} p(c_{\text{true}}, d_{l_t}|d_{l_1:l_{t-1}})\, P(c_{\text{true}}|d_{l_1:l_t})\, dd_{l_t} && (10)\\ &= \int \sum_{c_{\text{true}} \in \mathcal{C}} p(c_{\text{true}}, d_{l_t})\, \frac{P(c_{\text{true}}|d_{l_1:l_{t-1}})}{P(c_{\text{true}})}\, P(c_{\text{true}}|d_{l_1:l_t})\, dd_{l_t} && (11)\\ &\approx \frac{1}{M} \sum_{i=1}^{M} \frac{P(c^{(i)}|d_{l_1:l_{t-1}})}{P(c^{(i)})}\, P(c^{(i)}|d_{l_1:l_{t-1}}, d_{l_t}^{(i)}) && (12) \end{aligned}$$

with respect to C_{true} and D_{l_t} given the previous descriptors $d_{l_1:l_{t-1}}$. Because the training samples are assumed to represent i.i.d. samples from the joint distribution $p(c_{\text{true}}, d_{l_t})$, the score can be approximated by a Monte Carlo estimate computed over the training set in Eq. (12) where $c^{(i)}$ denotes the class of the i-th image in the training set, $d_{l_t}^{(i)}$ denotes the descriptor in the i-th training image at location l_t, and M denotes the total number of images in the training set. All the posterior probabilities can be obtained using Eq. (5).

Computing the Monte Carlo estimate can be time-consuming because all descriptors in the training set have to be considered. However, the NN distances required for the likelihoods can be computed in advance so that the overall score computation is still significantly faster than having to process all descriptors from the query image. In addition, it would be possible to only use a subset of the training samples where each sample would be drawn with a probability given by the current belief distribution.

For the special case where no descriptors have been extracted ($t = 1$) or where one chooses to ignore previously extracted descriptors, we can compute a score that ignores the current belief distribution and only maximizes the normalized expected likelihood (MEL). Plugging in $P(c^{(i)})$ for the current belief distribution in Eq. (12) results in

$$S_{\mathrm{MEL}}(l_t) = E[P(C_{\mathrm{true}}|D_{l_t})] \tag{13}$$

$$\approx \frac{1}{M}\sum_{i=1}^{M}\frac{P(c^{(i)})}{P(c^{(i)})}\,P(c^{(i)}|d_{l_t}^{(i)}) \tag{14}$$

$$= \frac{1}{M}\sum_{i=1}^{M}\frac{P(c^{(i)})\,p(d_{l_t}^{(i)}|c^{(i)})}{\sum_{c\in\mathcal{C}}P(c)\,p(d_{l_t}^{(i)}|c)}. \tag{15}$$

Because this score is independent of previous descriptors, it can be computed offline and is thus extremely fast.

3.2 Maximum Expected Information Gain

A popular method for feature selection is the maximum expected information gain (MIG) [18]. Here we consider a "dynamic" information gain version that takes previous descriptors into account during the recognition process [12,15]. It is given by the expected uncertainty/entropy reduction resulting from observing a new descriptor d_{l_t}. The information gain score S_{MIG} is the conditional expectation of this reduction with respect to D_{l_t} given the previous descriptors $d_{l_1:l_{t-1}}$:

$$S_{\mathrm{MIG}}(l_t) = H(C|d_{l_1:l_{t-1}}) - E[H(C|d_{l_1:l_{t-1}}, D_{l_t})|d_{l_1:l_{t-1}}] \tag{16}$$

$$= H(C|d_{l_1:l_{t-1}}) - \int\sum_{c_{\mathrm{true}}\in\mathcal{C}} p(c_{\mathrm{true}}, d_{l_t}|d_{l_1:l_{t-1}})H(C|d_{l_1:l_t})\,dd_{l_t} \tag{17}$$

$$\approx H(C|d_{l_1:l_{t-1}}) - \frac{1}{M}\sum_{i=1}^{M}\frac{P(c^{(i)}|d_{l_1:l_{t-1}})}{P(c^{(i)})}\,H(C|d_{l_1:l_{t-1}}, d_{l_t}^{(i)}) \tag{18}$$

with entropy

$$H(X) = -\sum_{x\in\mathcal{X}} P(x)\log P(x). \tag{19}$$

Like for S_{MEP}, the expected value is approximated by a Monte Carlo estimate using samples from the training set in Eq. (18). Note that the information gain is

independent of the true class, meaning that a high MIG score only requires the resulting posterior distribution to be "non-uniform", thus completely ignoring how probable the true class is.

3.3 Intrinsically Two-Dimensional Signals

The following image-based selection method uses a saliency operator which detects intrinsically two-dimensional ($I2D$) signals [19]. The intrinsic dimensionality of a signal $u(x, y)$ is defined as $I0D$ for all signals that are constant and as $I1D$ for all signals that can be written as a function of one variable in an appropriately rotated coordinate system (e.g. an image of an oriented straight edge). In contrast, $I2D$-signals make full use of the two degrees of freedom (e.g. an image of a corner or crossing lines). The $I2D$-saliency also appears to play an important role in the control of saccadic eye movements [5,16] which motivates its use as a score function within the context of this work. In order to identify the interesting $I2D$-points, we make use of the generalized curvature operator introduced in [19]: The generalized curvature operator $T_n : C^2(\Omega) \to C(\Omega)$ with compact $\Omega \subset \mathbb{R}^2$ is defined for $n \in \mathbb{N}$ by

$$T_n(u)(x) = \frac{1}{4}\left((\Delta u)^2 - \epsilon_n(u)^2\right) = \frac{1}{4}\underbrace{(\Delta u + |\epsilon_n(u)|)}_{=\lambda_1(u)}\underbrace{(\Delta u - |\epsilon_n(u)|)}_{=\lambda_2(u)} \tag{20}$$

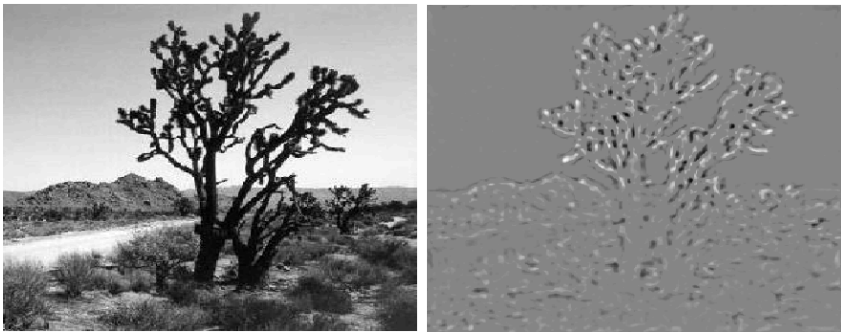
with eccentricity $\epsilon_n(u)^2 = (c_n * u)^2 + (s_n * u)^2$. The convolution kernels c_n and s_n are defined by their Fourier transform in polar coordinates ($x_1 = r\cos(\phi)$, $x_2 = r\sin(\phi)$) by

$$\mathcal{F}(c_n)(r, \phi) = (i)^n f(r)\cos(n\phi)$$
$$\text{and } \mathcal{F}(s_n)(r, \phi) = (i)^n f(r)\sin(n\phi).$$

f is a continuous function of the radius r given by $f(r) = 2\pi r^2 e^{\frac{1}{2}\frac{r^2}{\sigma_r^2}}$. λ_1 and λ_2 are the eigenvalues of the Hessian matrix of u in the case of $n = 2$ where the generalized curvature becomes the Gaussian curvature. The Gaussian curvature allows a distinction between elliptic, hyperbolic, and parabolic regions on the curved surface $\{(x, y, u(x, y))^T | (x, y)^T \in \mathbb{R}^2\}$. Using the eigenvalues, the clipped eigenvalue is defined by

$$CE(u) = |\min(0, \lambda_1(u))| - |\max(0, \lambda_2(u))|. \tag{21}$$

In contrast to directly using generalized curvature as a score function, the advantage of the clipped eigenvalue is that it can distinguish between positive elliptic and negative elliptic points, i.e., both eigenvalues are positive or negative. Furthermore, the clipped eigenvalue does not respond to hyperbolic regions. The latter is useful because hyperbolic regions are often found right next to elliptic ones, in which case the hyperbolic regions would only provide redundant



(a) Original (b) $I2D$-saliency

Fig. 1. Extracted $I2D$-saliency (b) of the image shown in (a). The extracted $I2D$-score is the clipped eigenvalue computed with the following parameters: $n = 6$, $\sigma_r = 0.2$. Positive elliptically curved regions are light and negative elliptically curved regions are dark.

information. The score function is then defined with respect to the luminance function u of the grid cell $\Omega(l_t)$ at location l_t by

$$S_{I2D}(l_t) = \frac{1}{|\Omega(l_t)|} \int_{\Omega(l_t)} |CE(u)(x)| \, dx. \tag{22}$$

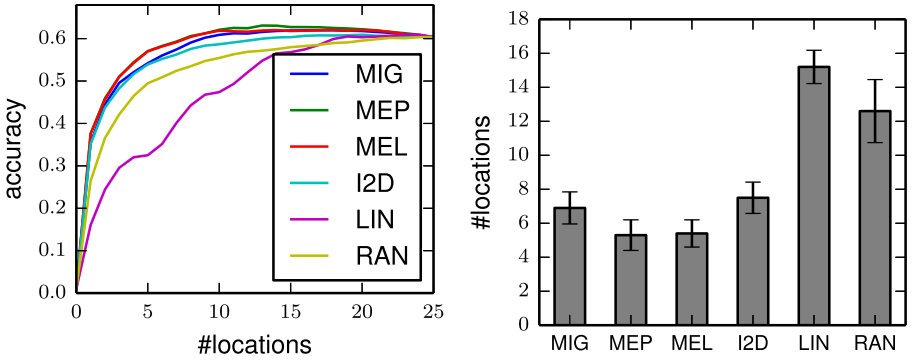
In contrast to belief-based score functions, the $I2D$-saliency is a purely image-based method. Consequently, it does not require any training data. The $I2D$-score function of an example image is illustrated in Fig. 1.

4 Evaluation

We evaluate the proposed information selection methods on the Caltech 101 data set [3]. We use 15 randomly selected images from each of the 101 object classes for training and 10 for testing. All images are scaled such that they have a maximum width or height of 300 pixels. Afterwards, densely-sampled SIFT descriptors are extracted (several thousands for each image depending on the size) and the NN distances are computed.[3]

Fig. 2 shows the mean accuracy over time for the different selection methods using a 5×5 grid and 10-fold cross validation. The MEP and MEL methods result in the quickest increase in accuracy and only require extracting descriptors from less than 6 grid cells on average for reliable classification (even though the MEL method ignores the current belief distribution). The MIG and I2D methods perform only slightly worse and all of the considered methods significantly outperform the baseline methods where descriptors are either selected randomly

[3] We use the code provided at https://github.com/sanchom/sjm for SIFT descriptor extraction and the FLANN library [10] for fast NN matches.



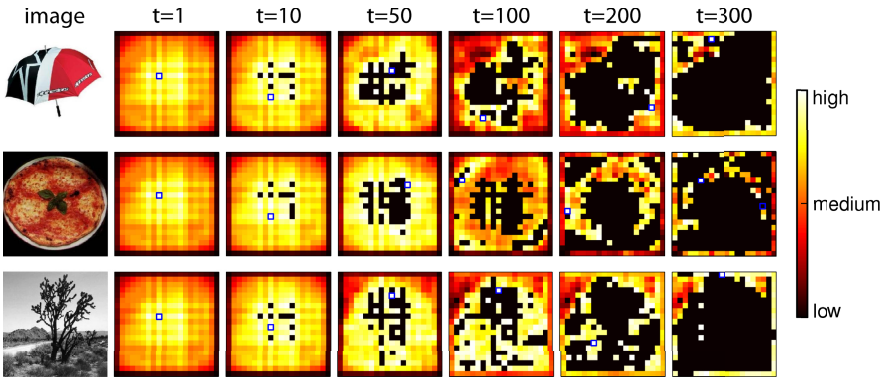
(a) accuracy (b) location count

Fig. 2. (a) Mean accuracy on the entire Caltech data set plotted for different time steps/location counts using different selection methods. (b) Mean number of time steps/location counts required for reaching at least 90% of the final accuracy where all descriptors have been extracted. The indicated standard deviation is computed with respect to the different folds.

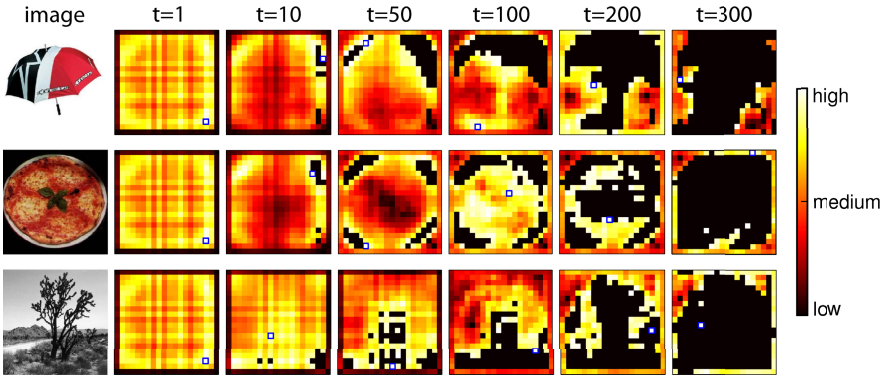
("RAN") or line by line starting at the top of the image ("LIN"). The final accuracy after having extracted *all* descriptors is identical for each method because the extraction order is irrelevant for the classification model. Interestingly, the accuracy is highest after having extracted about half of all descriptors (except for the baseline methods), showing that the remaining descriptors tend to only decrease the recognition performance.

To illustrate the process of sequentially selecting descriptors, Fig. 3 shows score distributions over time using a 20×20 grid for three example images. For the belief-based MEP and MIG selection methods shown in (a) and (b), the score distributions change significantly over time and adapt themselves to the query image based on the current belief distribution. The I2D score distribution remains constant over time aside from setting the score of previously selected locations to 0 (the apparent change in other locations is due to scaling in the visualization). At $t = 1$, both the MEP and the MIG scores are independent of the query image and only the I2D method uses the image information. Over time, the MEP and MIG scores adapt themselves to the current belief distribution over object classes, whereas the I2D score remains unchanged. The visible "grid pattern" (especially for $t \leq 10$) is an artifact resulting from some grid cells containing more descriptors than others (this could be avoided if all cells contained roughly the same number of descriptors).

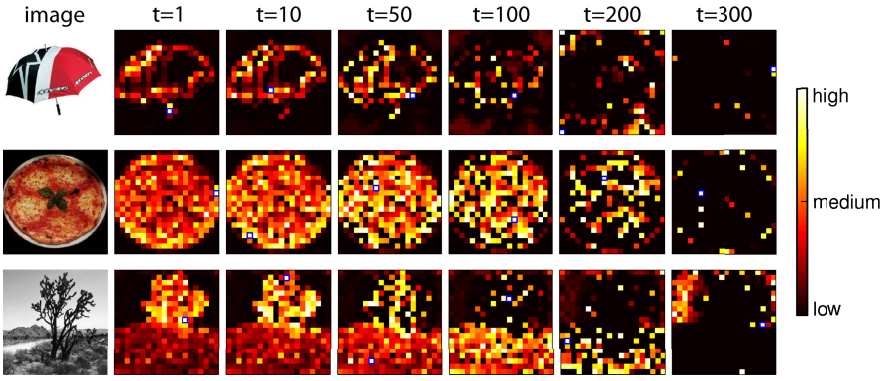
Perhaps surprisingly, the MEP score is highest at the center while the MIG score is initially highest in the periphery. One possible explanation for this effect is that the MEP method can be interpreted as a "confirmation strategy" whereas the MIG method can be interpreted as a "discriminative strategy". For MEP, extracting descriptors from the center of an object usually increases the probability of the true class without necessarily resulting in a unique classification



(a) MEP



(b) MIG



(c) I2D

Fig. 3. Examples of score distributions over time using a 20×20 grid for different selection methods and query images. The small blue square indicates the cell with the highest score from which the next descriptor(s) are extracted. Cells that have already been selected have a score value of 0 (black).

(i.e. the overall belief distribution can still be very uniform). In contrast, the MIG method is agnostic with respect to the true class and only seeks to reduce uncertainty (e.g. by ruling out large numbers of classes). This could be accomplished by analyzing the "context" of objects, which is why the MIG method might first focus on the background.

5 Conclusion

We have proposed different methods for adaptive information selection from images where the current belief distribution directly determines which image locations should be considered next. In addition, we have also considered an image-based selection method that does not require any training data. Using these methods, we have extended the NBNN approach and we have shown that the selection methods make it possible to only consider a small subset of the available information while maintaining the original recognition performance. In particular for NBNN, where computing the NN distances for each descriptor is very time-consuming, the result is a significantly reduced computation time.

One of the problems not addressed in this paper is the fact that features in close proximity to each other are highly correlated. While the naive Bayes assumption can be justified for inference by the greatly reduced computational complexity, for the information selection it would be possible to use a more sophisticated model where correlations are explicitly considered. As a result, there would be a penalty for extracting features located very closely to each other, thus avoiding processing of redundant information.

In this paper, we have considered belief-based selection strategies (MEP, MIG) and image-based strategies (I2D) separately. A more promising approach could be a combination of both strategies [16] because the belief-based strategy completely ignores what is readily available in the image while a purely image-based strategy has difficulties selecting the relevant information because it ignores the training data. Due to the complementary nature of these strategies, a hybrid strategy could further improve the selection process.

We believe that the proposed selection methods can also be useful for problems beyond recognizing single objects. Especially for complex scenes containing many objects, an adaptive information selection strategy could predict the likely locations of objects and thereby facilitate understanding of the entire scene.

Finally, the general nature of the proposed information selection approaches allows for the application to systems which must perform actions to obtain new information from their environments (e.g. an autonomous spacecraft [14] or a melting probe [13]). These actions can cause high costs in terms of, for example, energy consumption or execution time. In these situations, it is thus highly desirable to avoid non-informative actions by using adaptive selection strategies.

Acknowledgments. This work was supported by the German Federal Ministry for Economic Affairs and Energy (DLR project "KaNaRiA", funding no. 50 NA 1318, and DLR project "CAUSE", funding no. 50 NA 1505).

References

1. Boiman, O., Shechtman, E., Irani, M.: In defense of nearest-neighbor based image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
2. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV, vol. 1, pp. 1–2 (2004)
3. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. Computer Vision and Image Understanding **106**(1), 59–70 (2007); special issue on Generative Model Based Vision
4. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
5. Krieger, G., Rentschler, I., Hauske, G., Schill, K., Zetzsche, C.: Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. Spatial Vision **13**(2–3), 201–214 (2000)
6. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
7. Liu, H., Sun, J., Liu, L., Zhang, H.: Feature selection with dynamic mutual information. Pattern Recognition **42**(7), 1330–1339 (2009)
8. Lowe, D.G.: Object recognition from local scale-invariant features. In: The Proceedings of the Seventh IEEE International Conference on Computer vision, 1999, vol. 2, pp. 1150–1157. IEEE (1999)
9. McCann, S., Lowe, D.G.: Local naive Bayes nearest neighbor for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3650–3656. IEEE (2012)
10. Muja, M., Lowe, D.G.: Scalable nearest neighbor algorithms for high dimensional data. IEEE Transactions on Pattern Analysis and Machine Intelligence **36** (2014)
11. Najemnik, J., Geisler, W.S.: Optimal eye movement strategies in visual search. Nature **434**(7031), 387–391 (2005)
12. Kluth, T., Reineking, T., Nakath, D., Zetzsche, C., Schill, K.: Active sensorimotor object recognition in three-dimensional space. In: Freksa, C., Nebel, B., Hegarty, M., Barkowsky, T. (eds.) Spatial Cognition 2014. LNCS, vol. 8684, pp. 312–324. Springer, Heidelberg (2014)
13. Niedermeier, H., Clemens, J., Kowalski, J., Macht, S., Heinen, D., Hoffmann, R., Linder, P.: Navigation system for a research ice probe for antarctic glaciers. In: IEEE/ION PLANS 2014, pp. 959–975. IEEE (2014)
14. Pavone, M., Acikmese, B., Nesnas, I.A., Starek, J.: Spacecraft autonomy challenges for next generation space missions (2013), online (to appear in Lecture Notes in Control and Information Systems). http://goo.gl/nU8xG0
15. Reineking, T., Schill, K.: Evidential object recognition based on information gain maximization. In: Cuzzolin, F. (ed.) BELIEF 2014. LNCS, vol. 8764, pp. 227–236. Springer, Heidelberg (2014)
16. Schill, K., Umkehrer, E., Beinlich, S., Krieger, G., Zetzsche, C.: Scene analysis with saccadic eye movements: Top-down and bottom-up modeling. Journal of Electronic Imaging **10**(1), 152–160 (2001)

17. Torralba, A., Oliva, A., Castelhano, M.S., Henderson, J.M.: Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. Psychological Review **113**(4), 766 (2006)
18. Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: ICML, vol. 97, pp. 412–420 (1997)
19. Zetzsche, C., Barth, E.: Image surface predicates and the neural encoding of two-dimensional signal variations. In: SC-DL tentative, pp. 160–177. International Society for Optics and Photonics (1990)